



Loss Functions for Medical Image Segmentation: A Taxonomy

Jun Ma

junma@njust.edu.cn

School of Science

Nanjing University of Science and Technology

Nanjing, China



Outline

- 1. Machine Learning Recipe**
- 2. Loss Overview**
- 3. Code & References**



1. The Recipe for Machine Learning

- ① Collect dataset
 - ② Define data representation (e.g., CNN architecture)
 - ③ Define a loss measuring performance (loss function)
 - ④ Minimize the loss (optimizer)
-
- Loss functions are one of the important ingredients in deep learning-based medical image segmentation methods.
 - We present a systematic taxonomy to sort existing loss functions into four meaningful categories. This helps to reveal links and fundamental similarities between them.



2. Loss Overview

Background

Over the past five years, various loss functions have been proposed for deep learning-based medical image segmentation.

Goal

In the following slides, I will present the loss functions in a chronological order, but sort them into four organized groups.

- Distribution-based loss
- Region-based loss
- Boundary-based loss
- Compound loss



2. Loss Overview

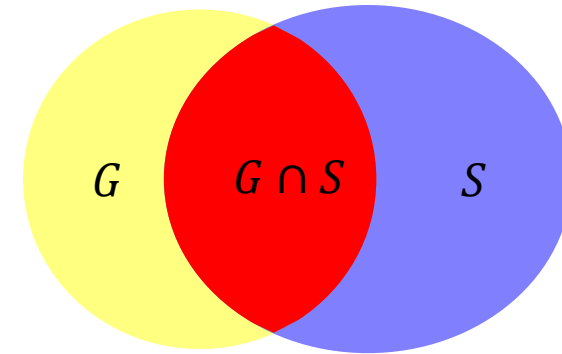
Two commonly used loss functions

Cross Entropy (CE)

$$D_{KL}(p \parallel q) = \underbrace{H(p, q)}_{\text{Cross entropy}} - \underbrace{H(p)}_{\text{Entropy}}$$

Distribution-based Loss

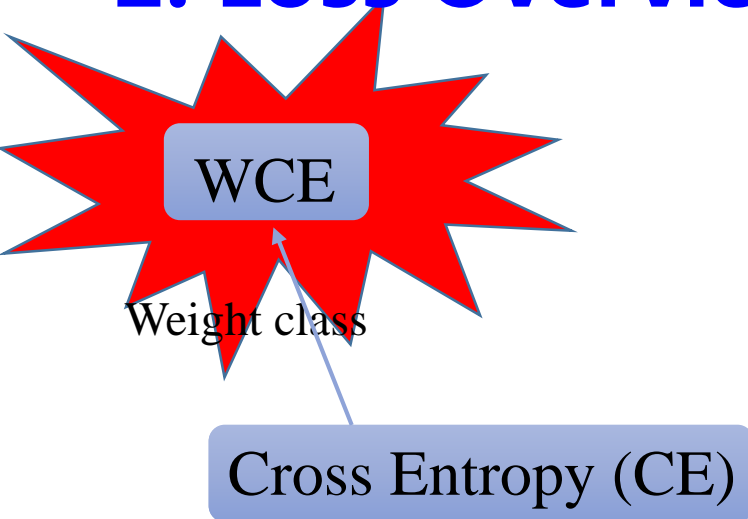
Dice



$$\text{Dice loss} = 1 - \frac{2|G \cap S|}{|G| + |S|}$$

Region-based Loss

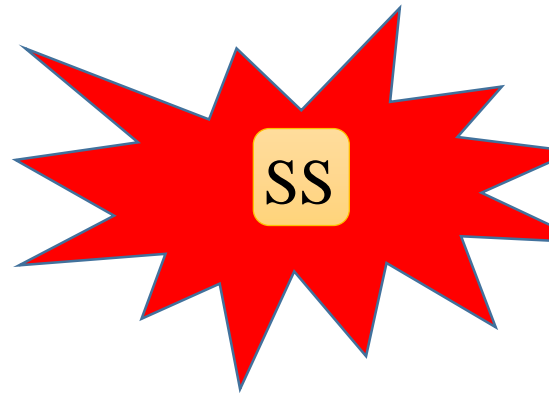
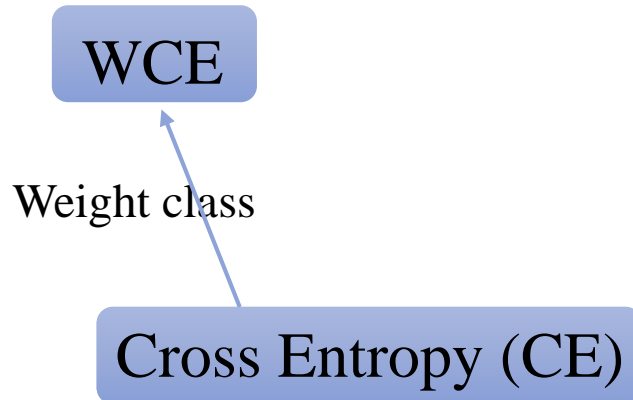
2. Loss Overview



$$L_{WCE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C w_c g_i^c \log s_i^c$$

Distribution-based Loss

2. Loss Overview



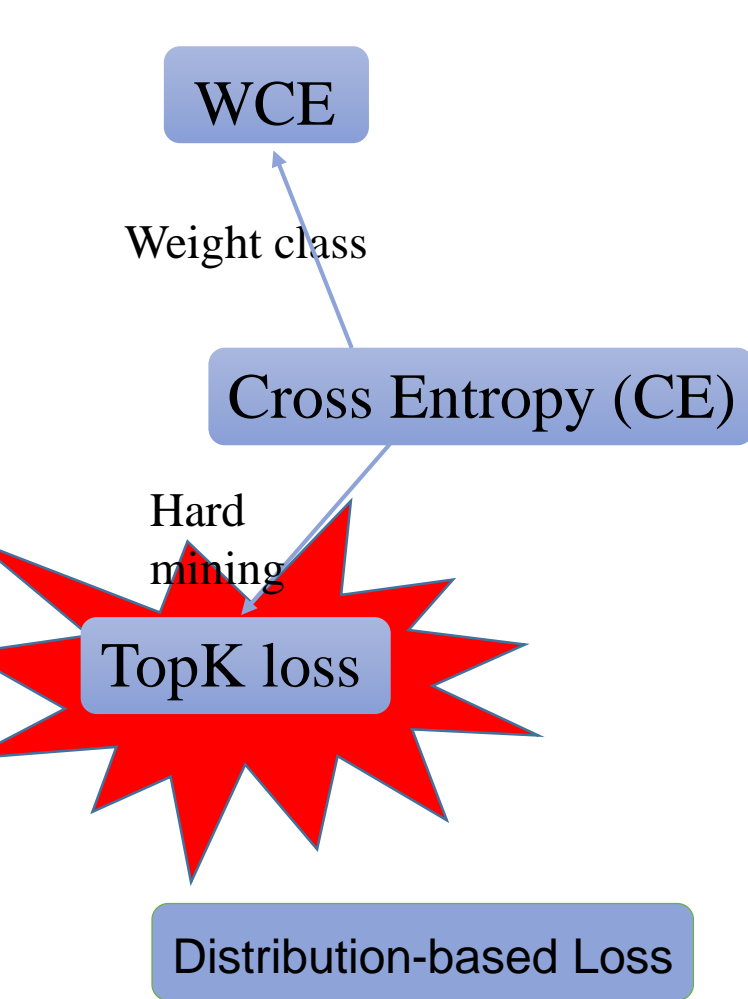
$$L_{SS} = w \frac{\sum_{i=1}^N \sum_c^C (g_i^c - s_i^c)^2 g_i^c}{\sum_{i=1}^N \sum_c^C g_i^c + \epsilon} + (1 - w) \frac{\sum_{i=1}^N \sum_c^C (g_i^c - s_i^c)^2 (1 - g_i^c)}{\sum_{i=1}^N \sum_{c=1}^C (1 - g_i^c) + \epsilon}$$

Weighted sum of the mean squared difference of **sensitivity** and **specificity**.

Distribution-based Loss

Region-based Loss

2. Loss Overview

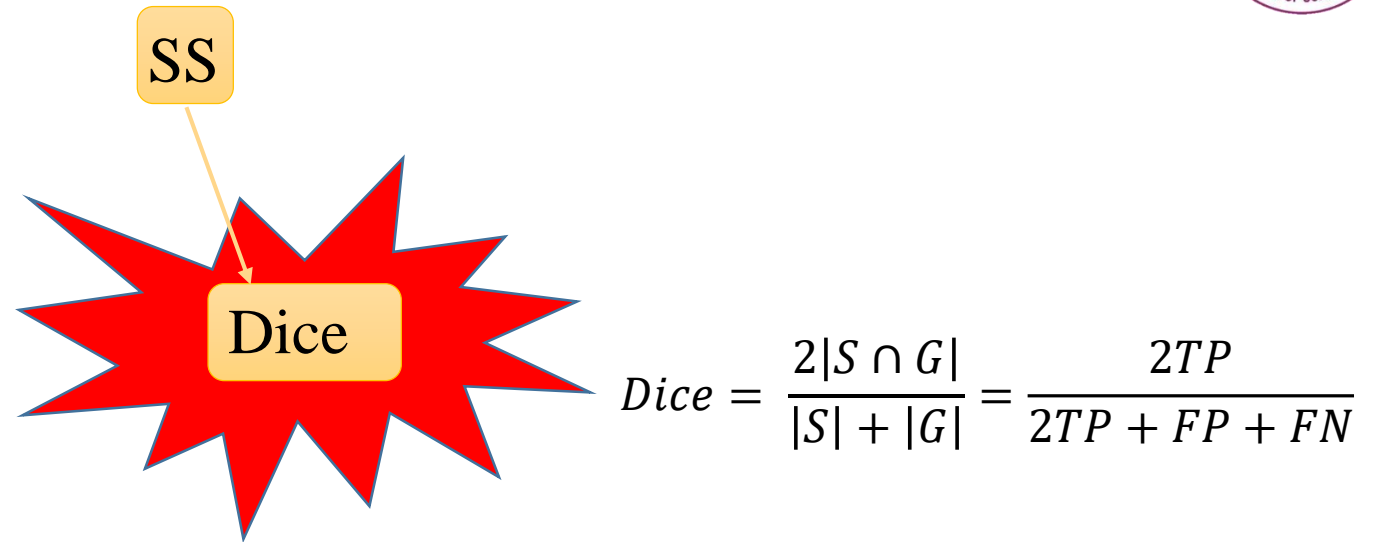
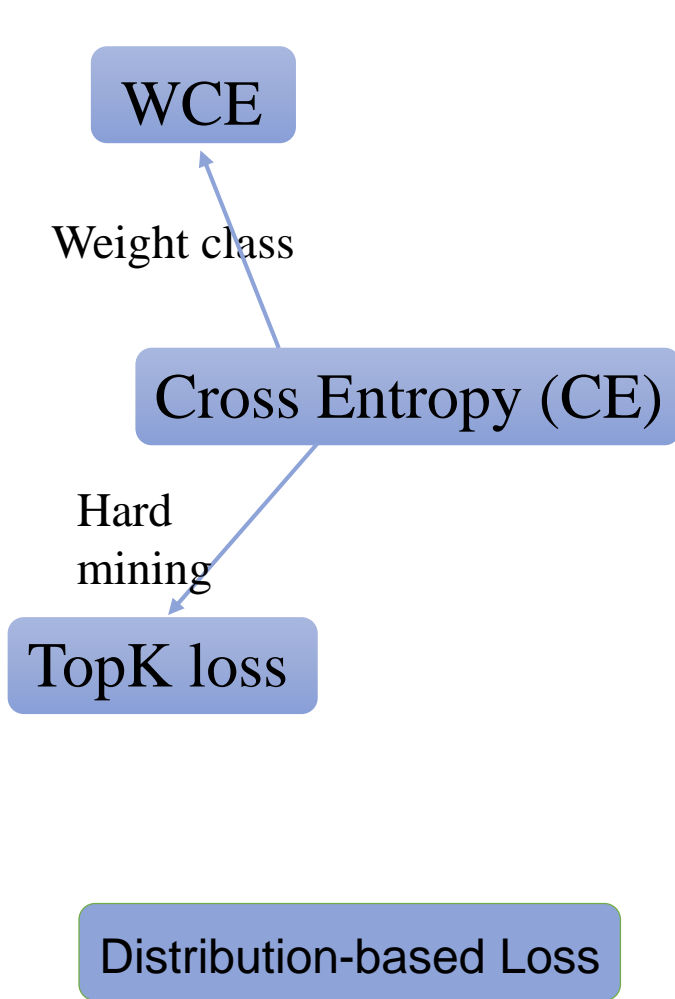


SS

$$L_{TopK} = - \frac{\sum_{i=1}^N \sum_{c=1}^C 1\{g_i = c \text{ and } s_i^c < t\} \log s_i^c}{\sum_{i=1}^N \sum_{c=1}^C 1\{g_i = c \text{ and } s_i^c < t\}}$$

Region-based Loss

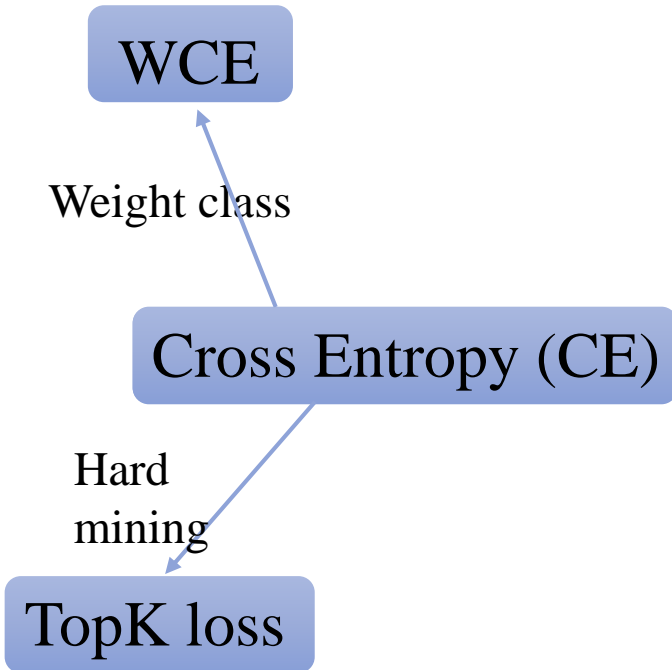
2. Loss Overview



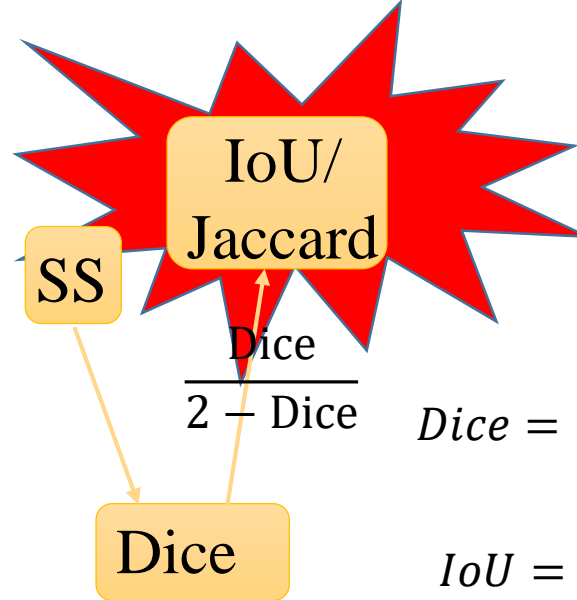
$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N \sum_{c=1}^C g_i^c s_i^c}{\sum_{i=1}^N \sum_{c=1}^C g_i^{c2} + \sum_{i=1}^N \sum_{c=1}^C s_i^{c2}}$$

Region-based Loss

2. Loss Overview



Distribution-based Loss



$$Dice = \frac{2|S \cap G|}{|S| + |G|} = \frac{2TP}{2TP + FP + FN}$$

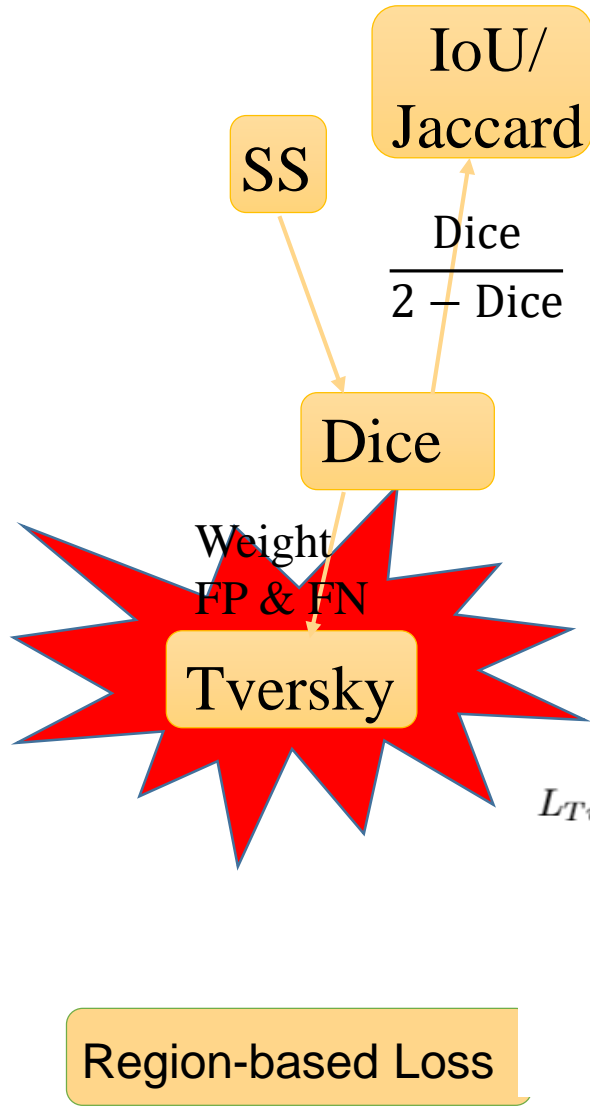
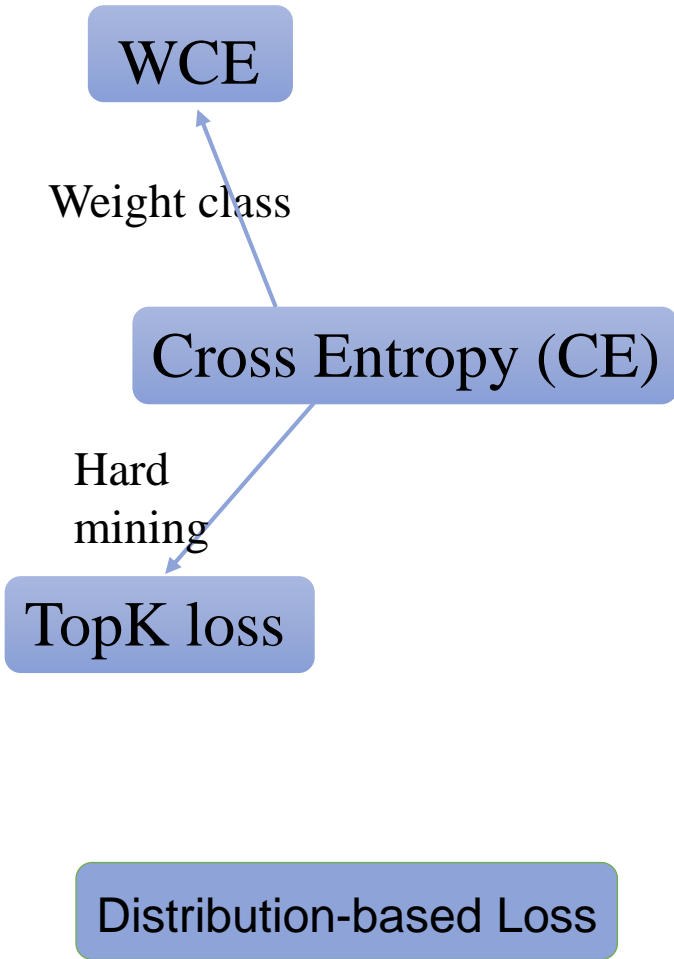
$$IoU = \frac{|S \cap G|}{|S \cup G|} = \frac{TP}{TP + FP + FN}$$

$$L_{IoU} = 1 - \frac{\sum_{i=1}^N \sum_{c=1}^C g_i^c s_i^c}{\sum_{i=1}^N \sum_{c=1}^C (g_i^c + s_i^c - g_i^c s_i^c)}$$

Region-based Loss



2. Loss Overview



$$Tversky = \frac{TP}{TP + \alpha FP + \beta FN}$$

$$L_{Tversky} = 1 - T(\alpha, \beta)$$

$$= 1 - \left(\frac{\sum_{i=1}^N \sum_{c=1}^C g_i^c s_i^c}{\sum_{i=1}^N \sum_{c=1}^C g_i^c s_i^c + \alpha \sum_{i=1}^N \sum_{c=1}^C (1 - g_i^c) s_i^c + \beta \sum_{i=1}^N \sum_{c=1}^C g_i^c (1 - s_i^c)} \right)$$

2. Loss Overview

IoU/



whats the difference with dice ? #17

Closed

argman opened this issue on 25 Nov 2018 · 1 comment



argman commented on 25 Nov 2018



Thanks for this great work, but i cannot understand, I think dice is also kind of optimizing iou, so whats the difference ?



bermanmaxim commented on 8 Jan

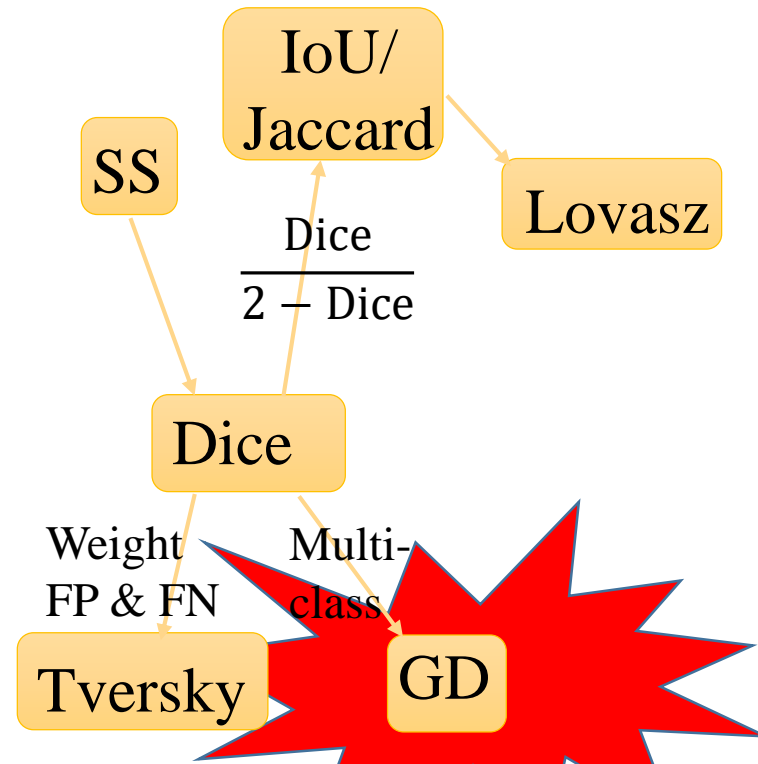
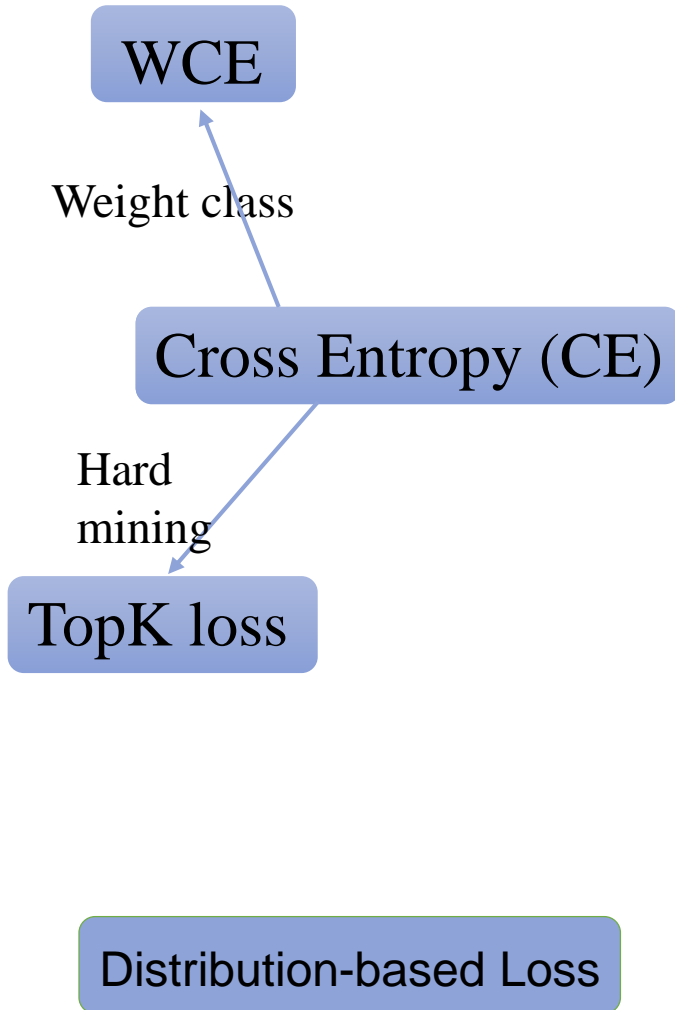
Owner



Dice is also a measure of performance based on discrete predictions. I think you are referring to "soft-Dice", often used to optimize dice. In practice we found that these naïve continuous versions of the discrete loss perform worse than our surrogate.



2. Loss Overview



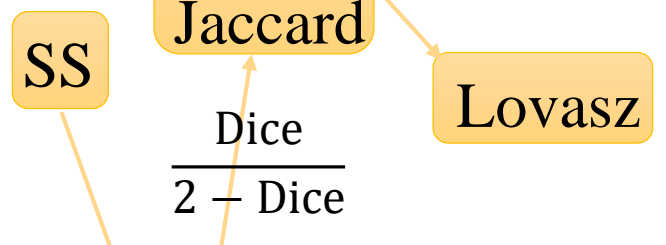
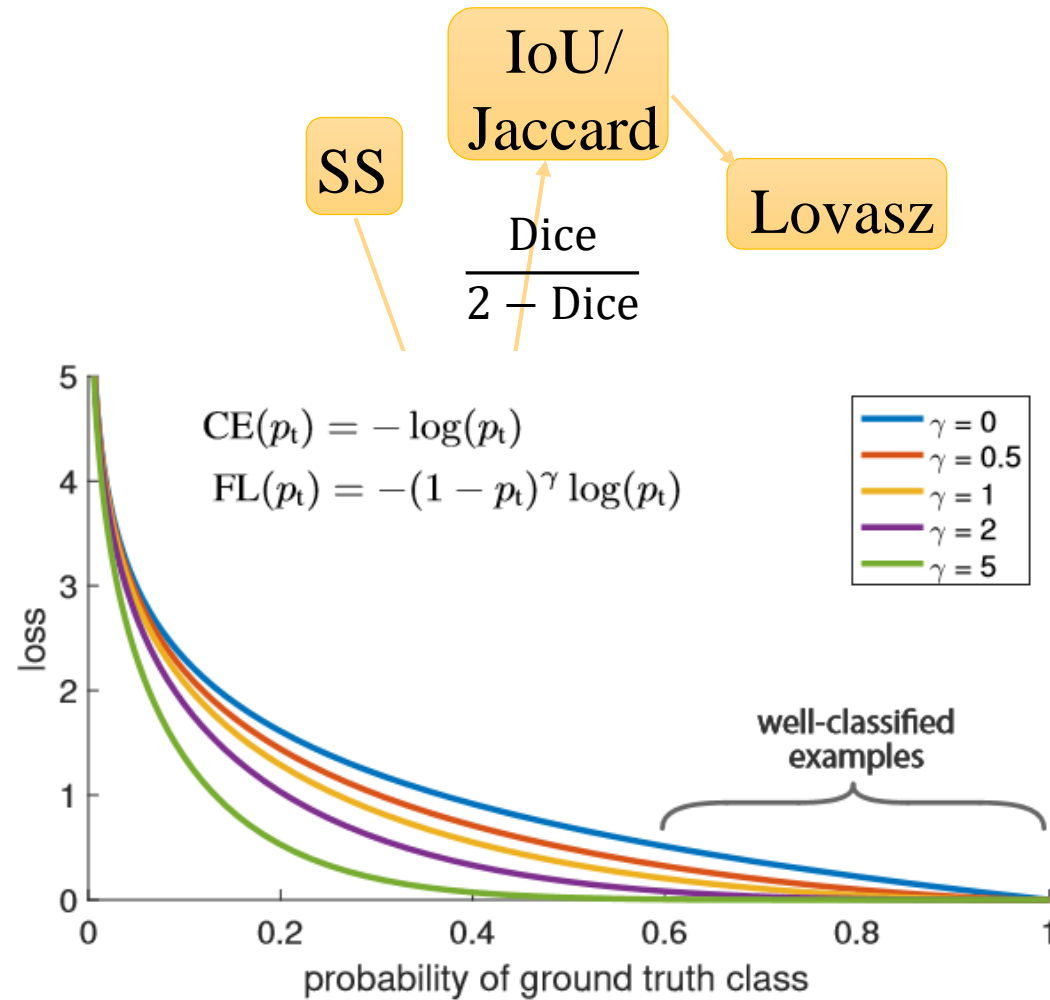
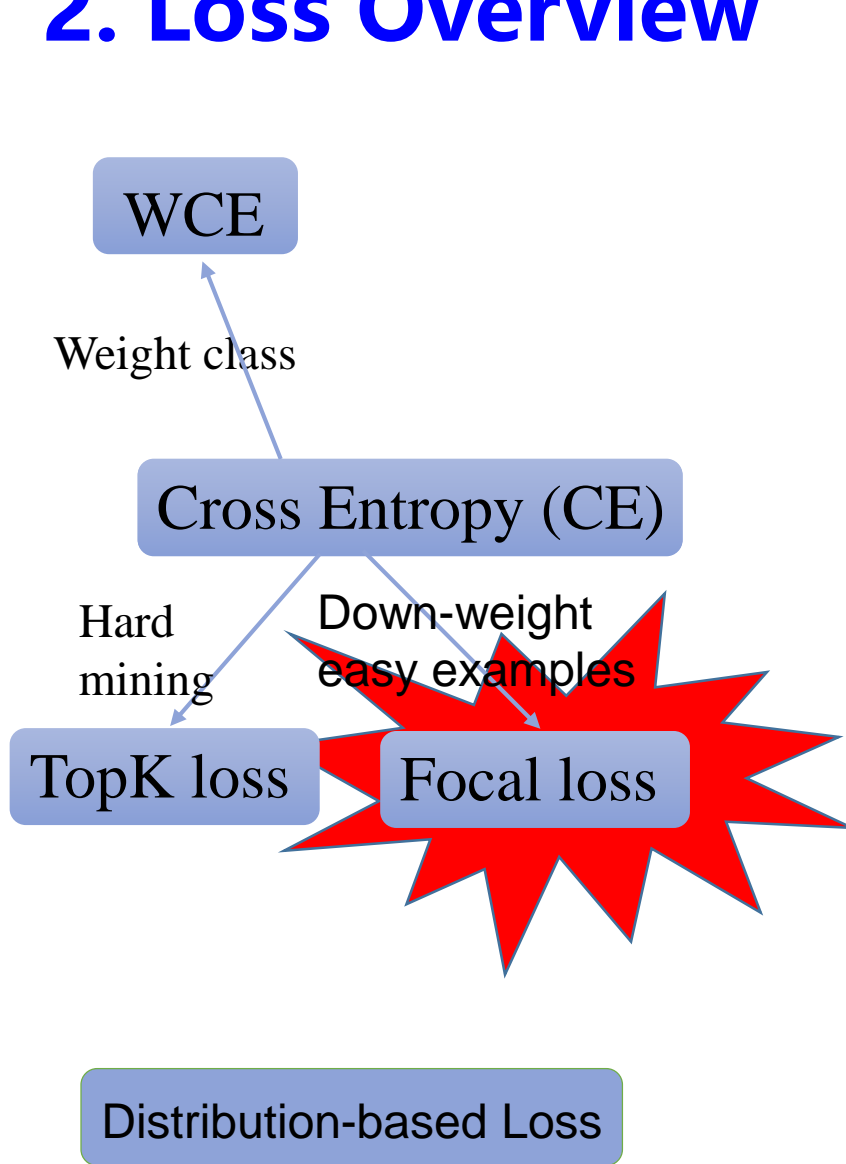
$$L_{GD} = 1 - 2 \frac{\sum_{c=1}^C w_c \sum_{i=1}^N g_i^c s_i^c}{\sum_{c=1}^C w_c \sum_{i=1}^N (g_i^c + s_i^c)} \quad (13)$$

where $w_c = \frac{1}{(\sum_{i=1}^N g_i^c)^2}$ is used to provide invariance to different label set properties.

Region-based Loss

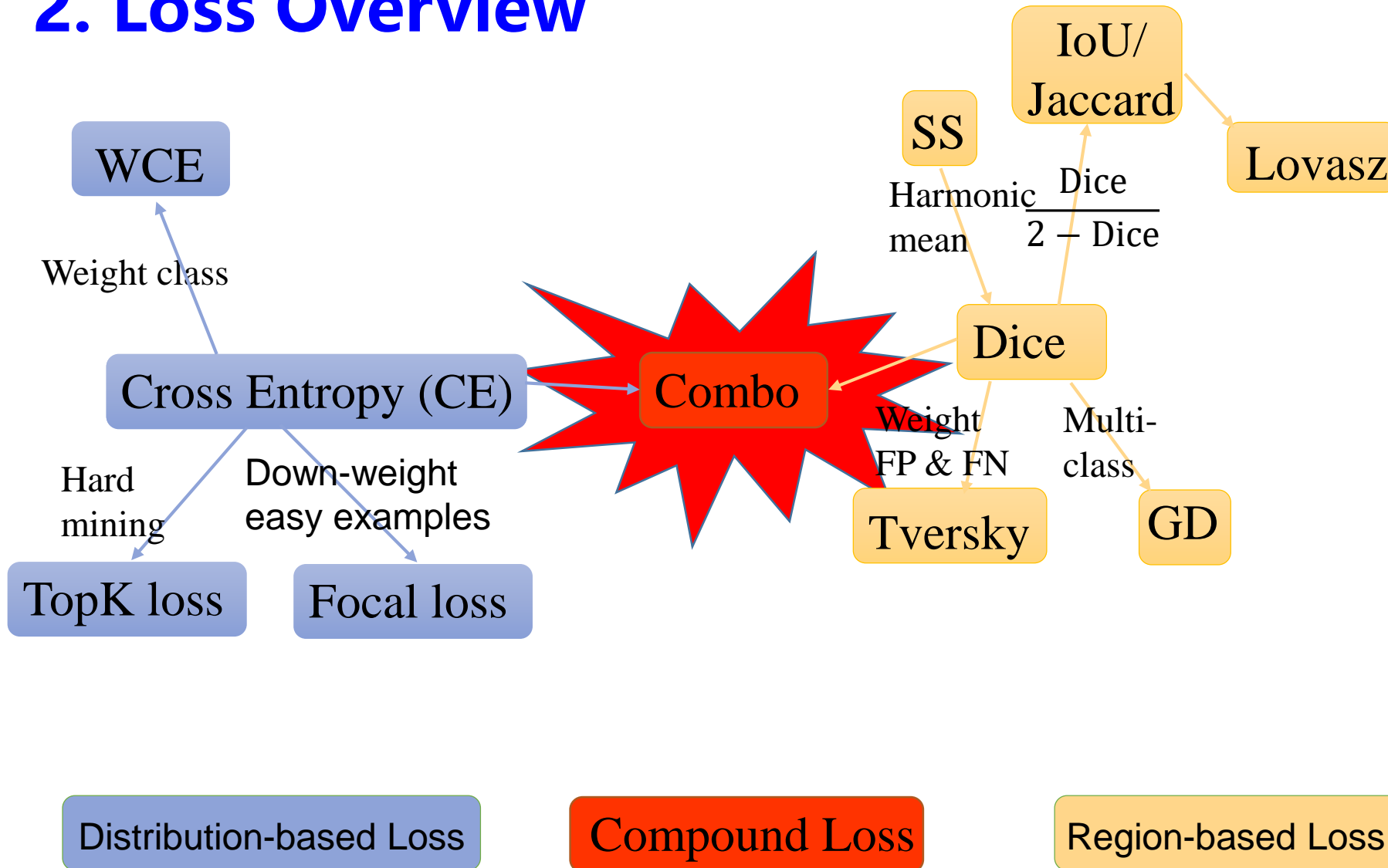


2. Loss Overview





2. Loss Overview

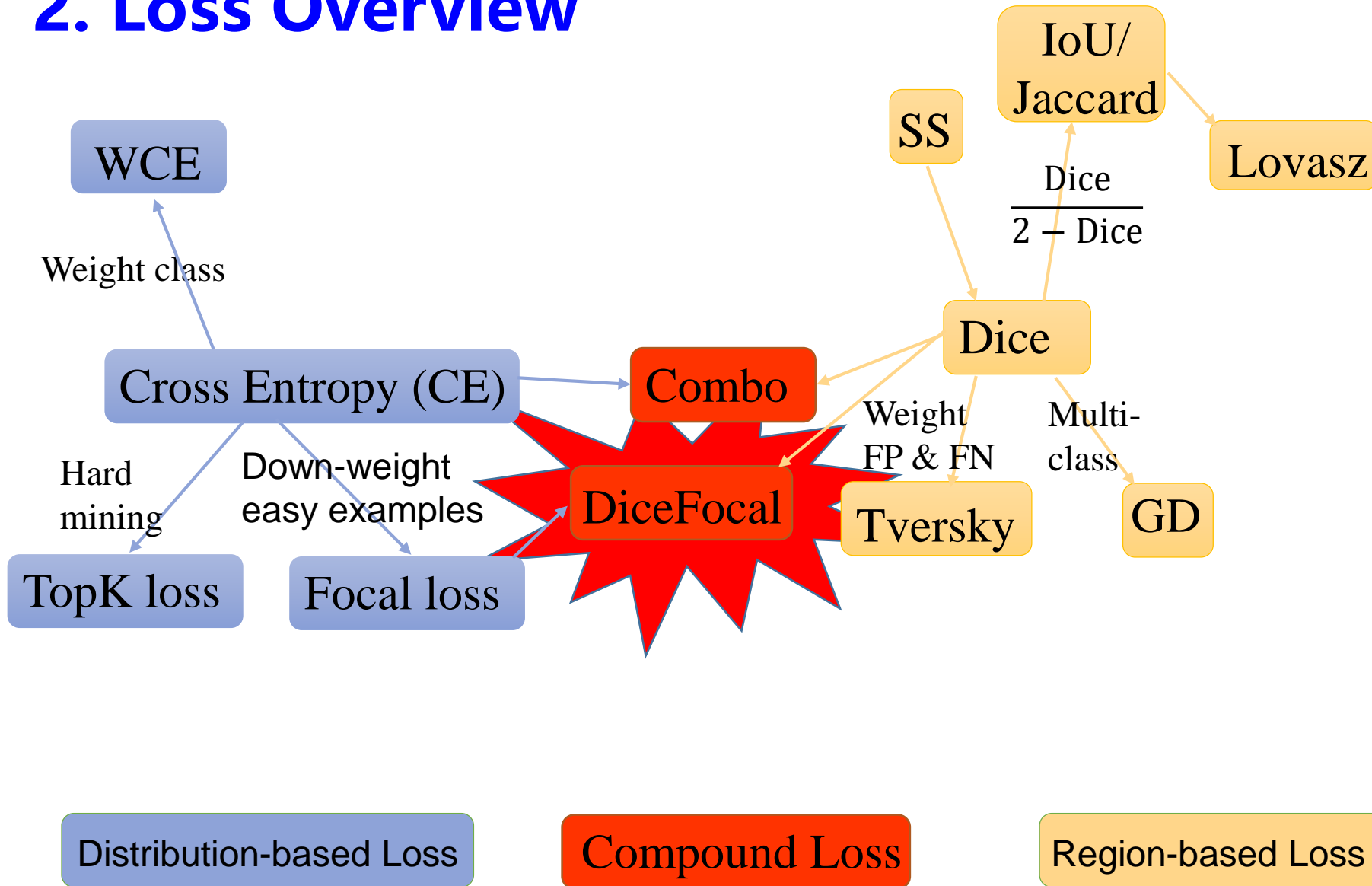


Taghanaki, S. A., Zheng, Y., Zhou, S. K., Georgescu, B., Sharma, P., Xu, D., ... & Hamarneh, G. "Combo loss: Handling input and output imbalance in multi-organ segmentation." *Computerized Medical Imaging and Graphics* 75 (2019): 24-33.

Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. "nnu-net: Self-adapting framework for u-net-based medical image segmentation." *arXiv preprint arXiv:1809.10486* (2018).

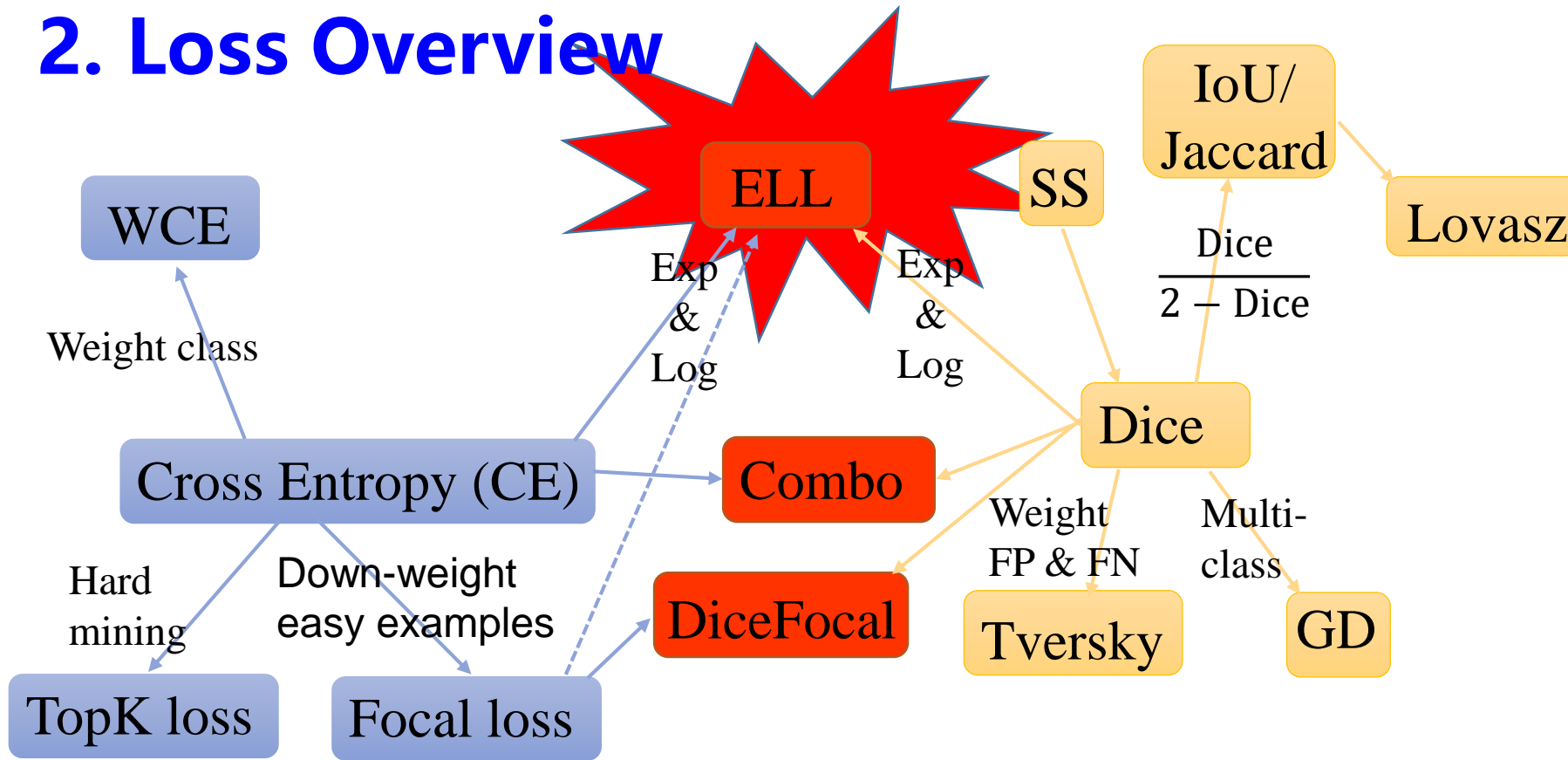


2. Loss Overview





2. Loss Overview



$$L_{ELL} = w_{Dice} E[(-\log(Dice_c))^{\gamma^{Dice}}] + w_{CE} E[w_c(-\log(s_i^c))^{\gamma^{CE}}]$$

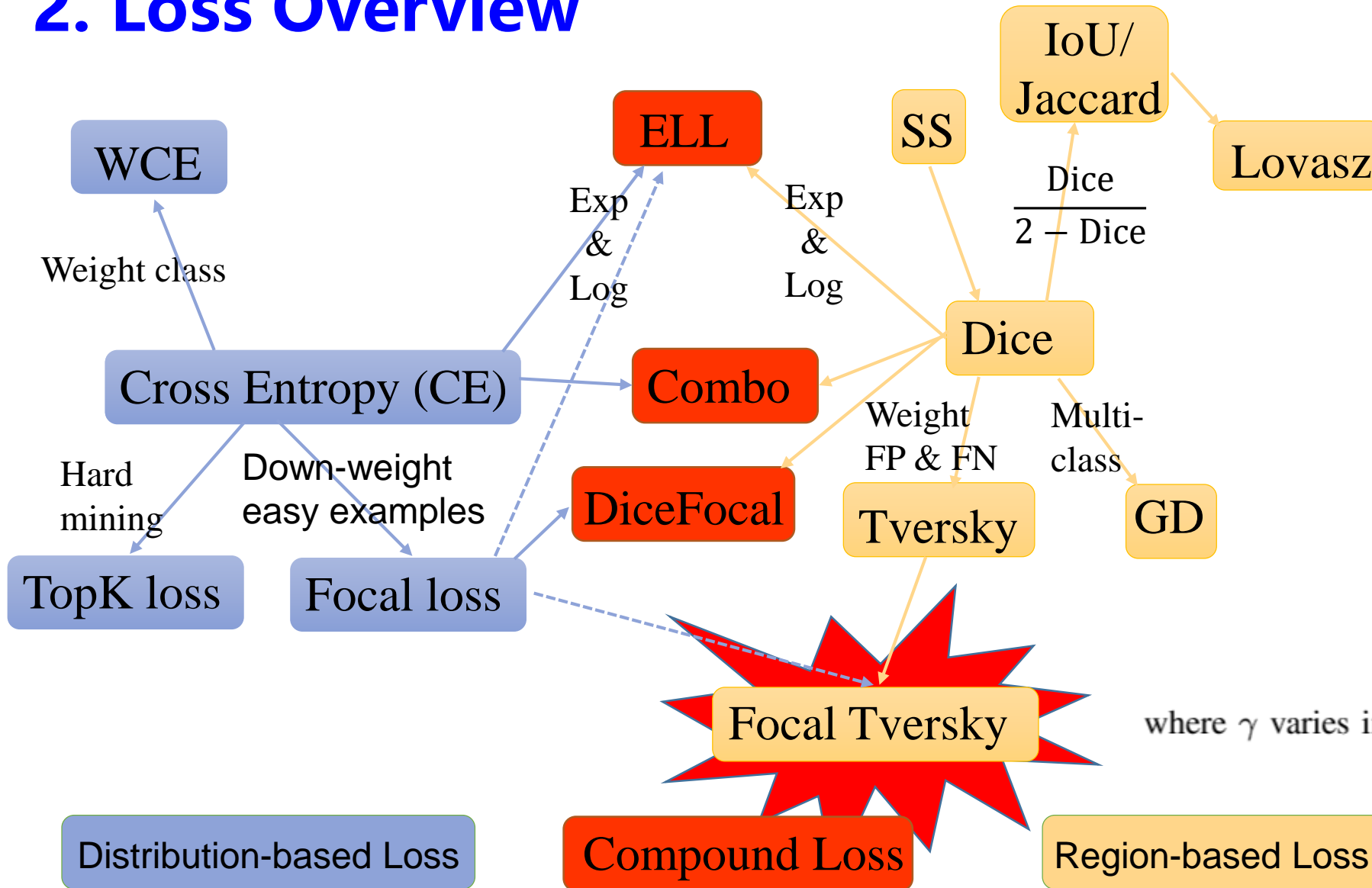
Distribution-based Loss

Compound Loss

Region-based Loss



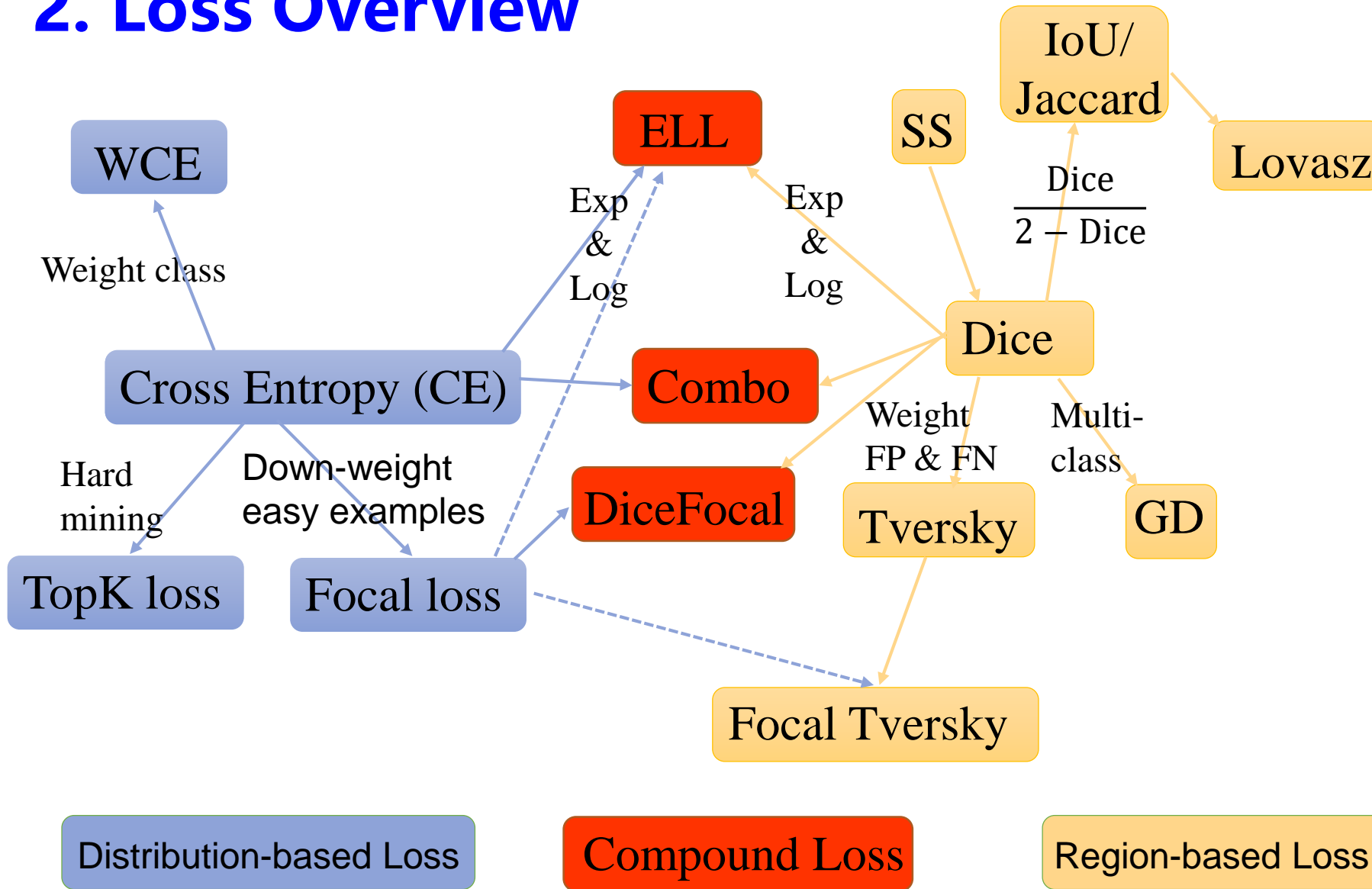
2. Loss Overview



$$L_{FTL} = (L_{Tversky})^{\frac{1}{\gamma}}$$

where γ varies in the range [1, 3].

2. Loss Overview



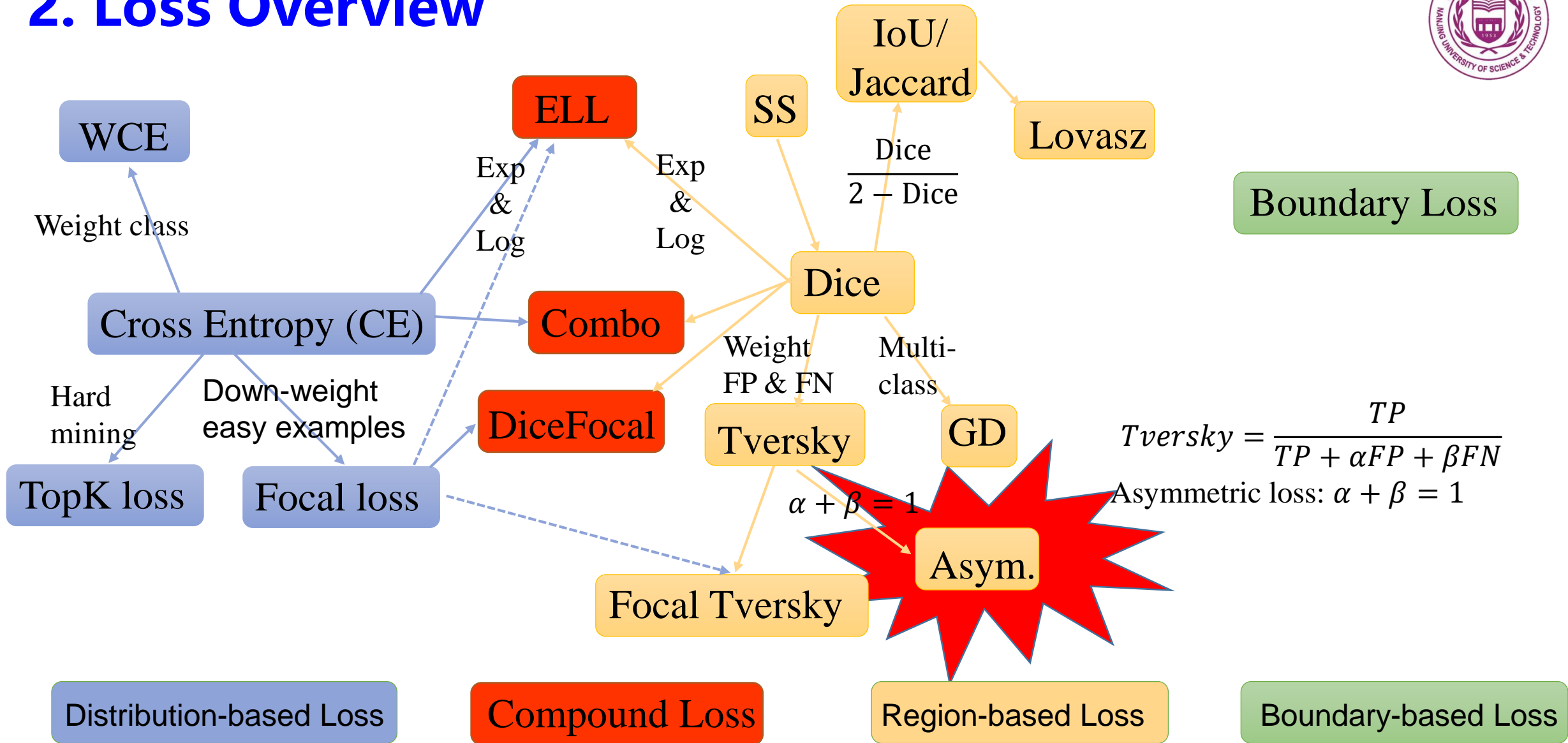
Boundary Loss

$$D_G(q) = \|q - z_{\partial G}(q)\|$$

$$\text{Dist}(\partial G, \partial S) = 2 \int_{\Delta S} D_G(q) dq$$

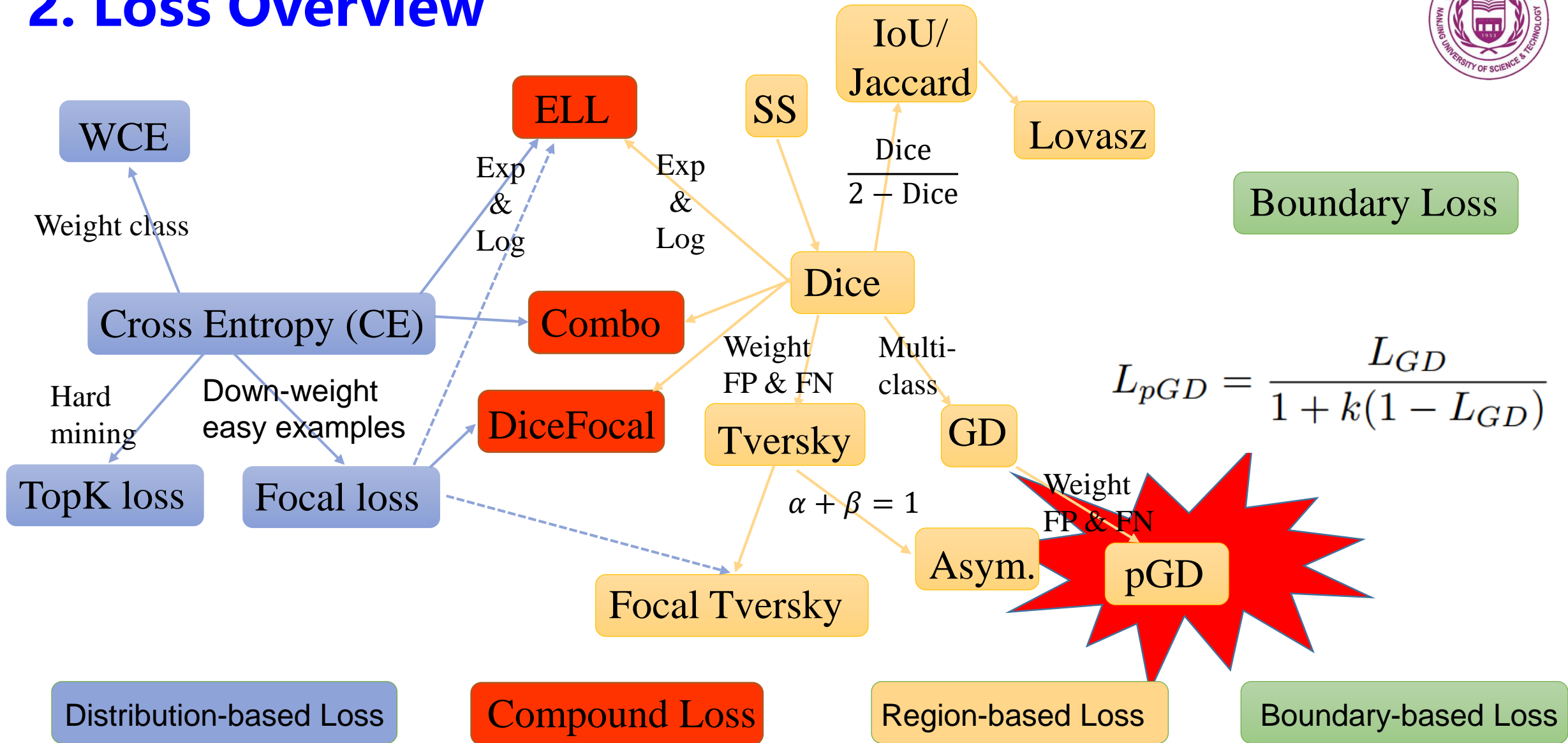


2. Loss Overview



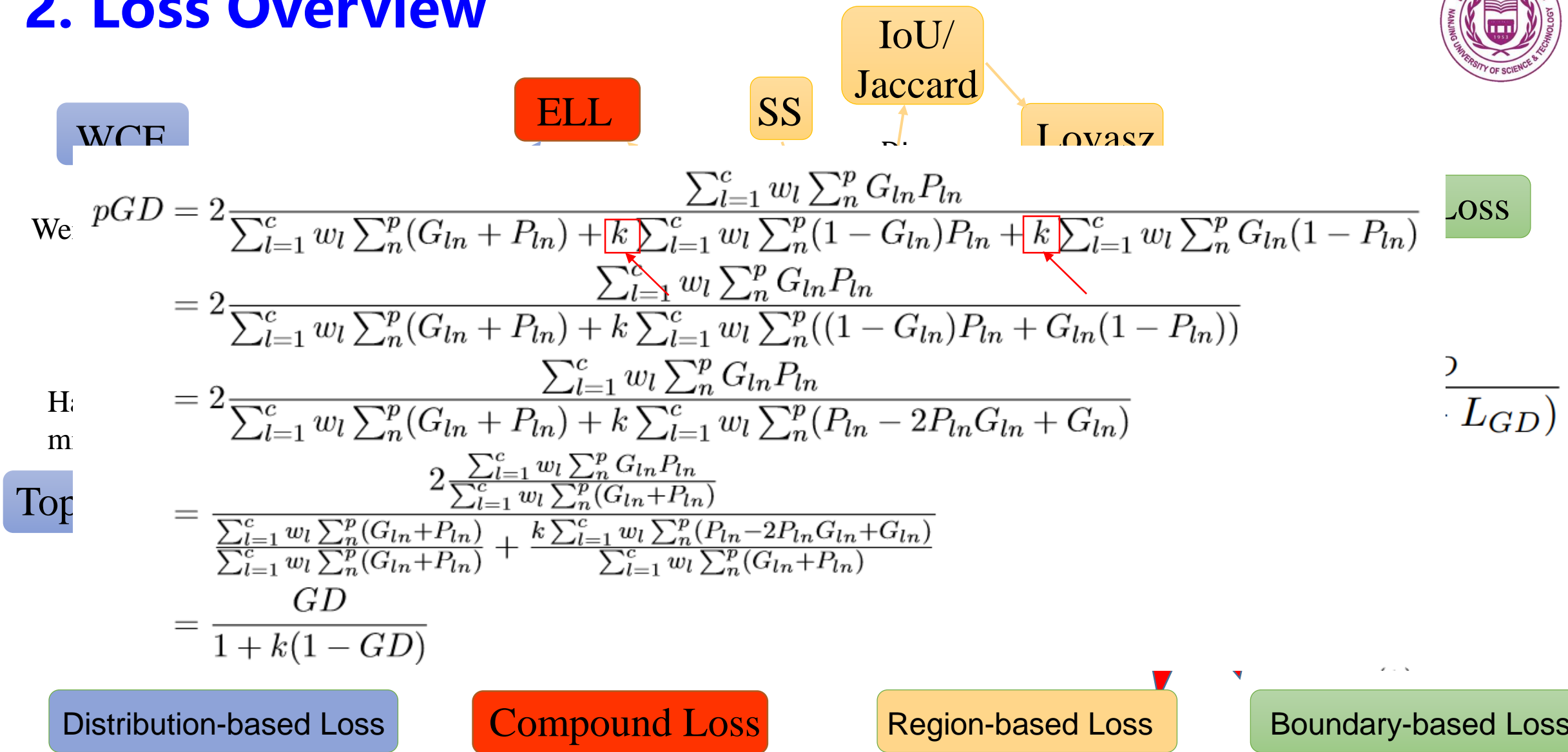


2. Loss Overview





2. Loss Overview



WCE

ELL

SS

IoU/
Jaccard

Lovasz

GD

$$\begin{aligned}
 \text{We } pGD &= 2 \frac{\sum_{l=1}^c w_l \sum_n^p G_{ln} P_{ln}}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln}) + k \sum_{l=1}^c w_l \sum_n^p (1 - G_{ln}) P_{ln} + k \sum_{l=1}^c w_l \sum_n^p G_{ln} (1 - P_{ln})} \\
 &= 2 \frac{\sum_{l=1}^c w_l \sum_n^p G_{ln} P_{ln}}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln}) + k \sum_{l=1}^c w_l \sum_n^p ((1 - G_{ln}) P_{ln} + G_{ln} (1 - P_{ln}))} \\
 \text{H: } &= 2 \frac{\sum_{l=1}^c w_l \sum_n^p G_{ln} P_{ln}}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln}) + k \sum_{l=1}^c w_l \sum_n^p (P_{ln} - 2P_{ln} G_{ln} + G_{ln})} \\
 \text{m } &= \frac{2 \frac{\sum_{l=1}^c w_l \sum_n^p G_{ln} P_{ln}}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln})}}{\frac{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln})}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln})} + \frac{k \sum_{l=1}^c w_l \sum_n^p (P_{ln} - 2P_{ln} G_{ln} + G_{ln})}{\sum_{l=1}^c w_l \sum_n^p (G_{ln} + P_{ln})}} \\
 \text{Top } &= \frac{GD}{1 + k(1 - GD)}
 \end{aligned}$$

Distribution-based Loss

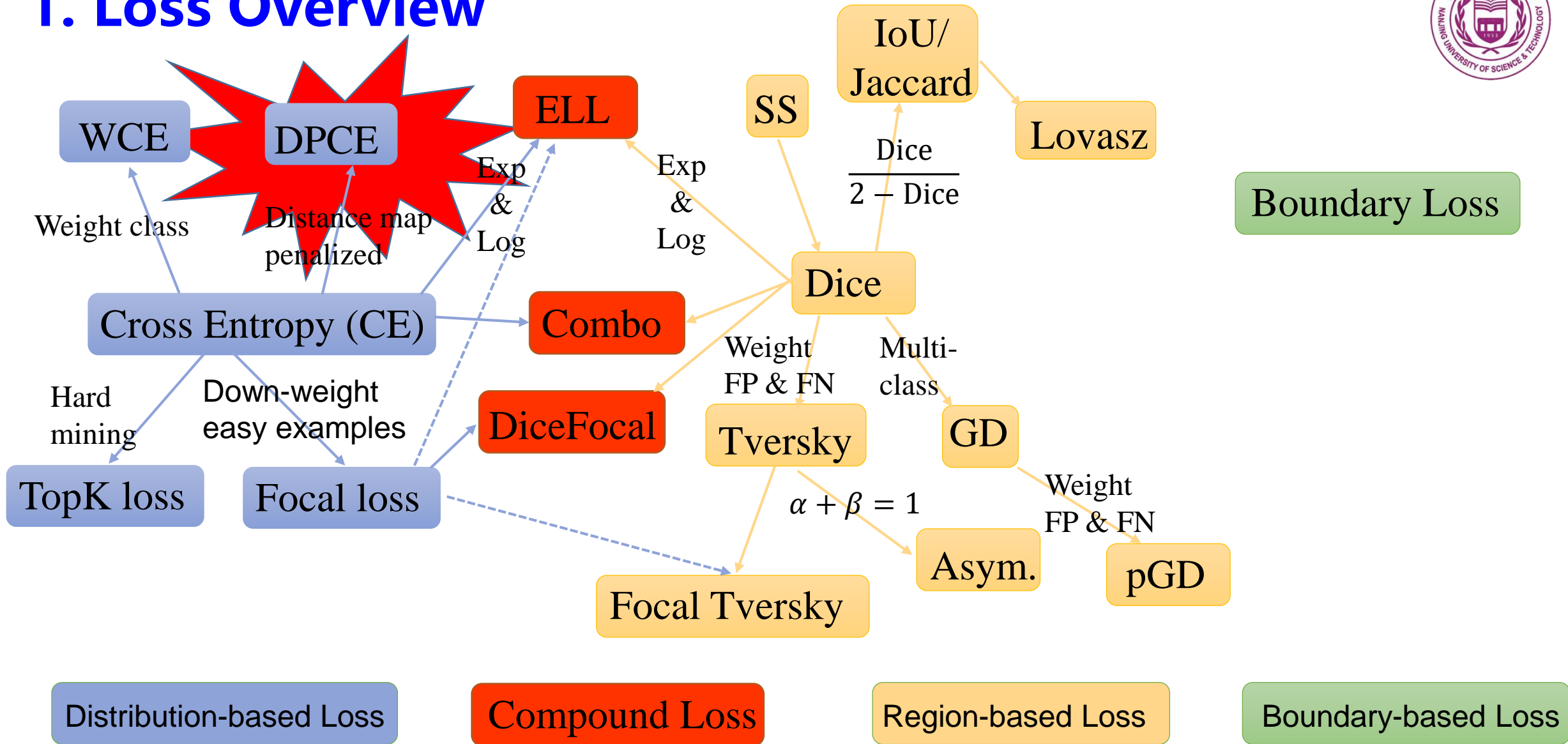
Compound Loss

Region-based Loss

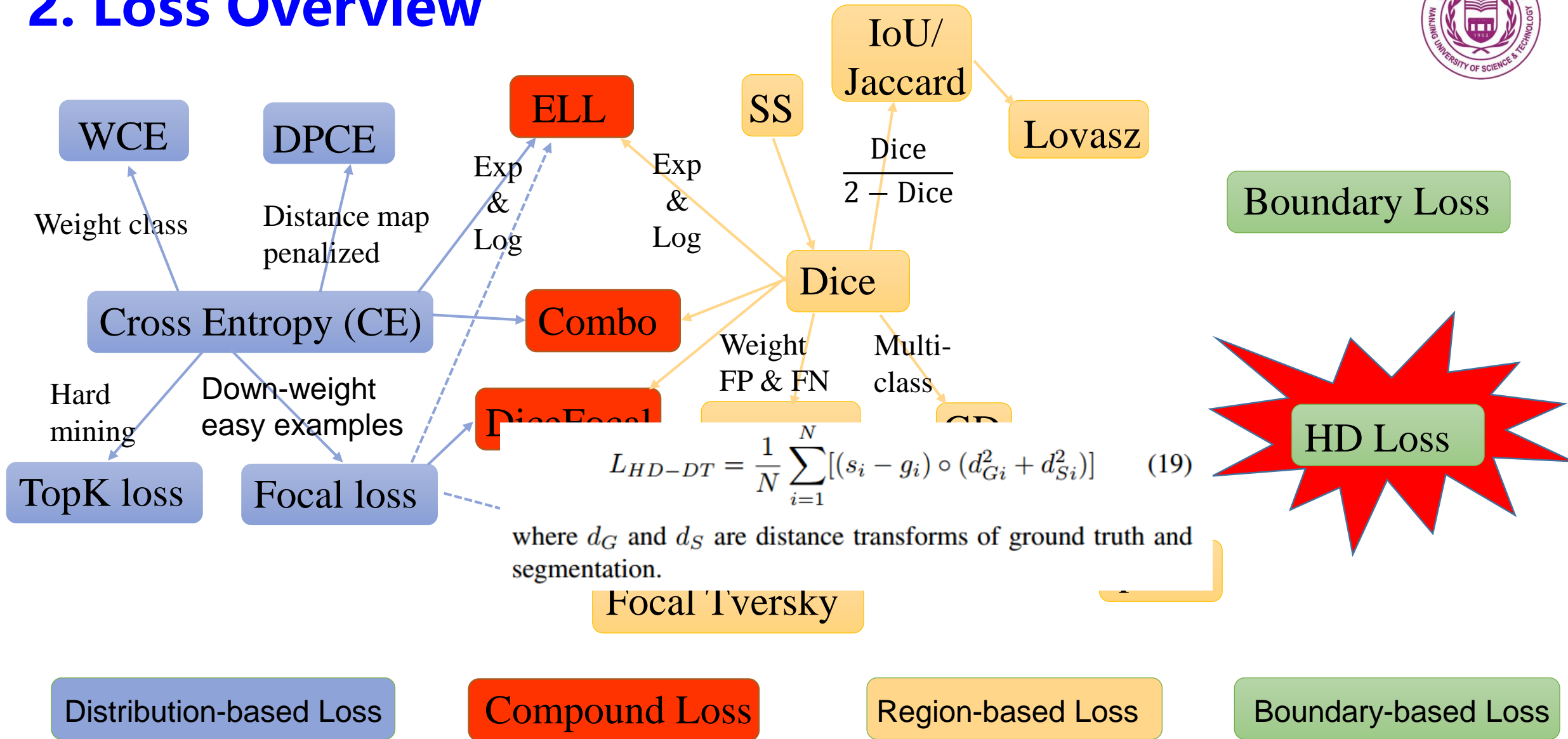
Boundary-based Loss



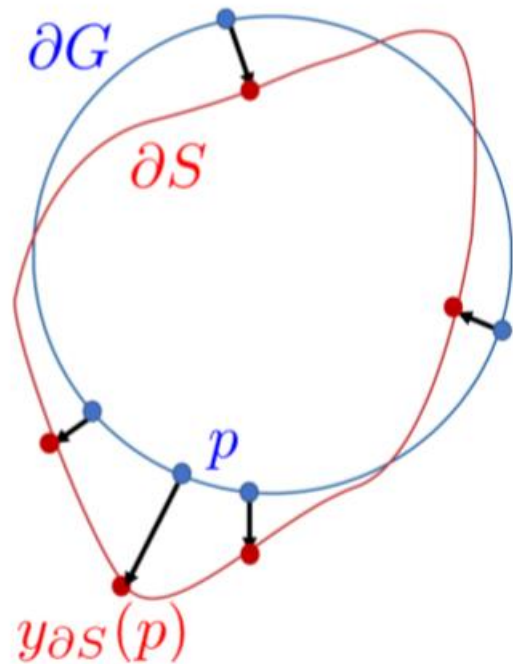
1. Loss Overview



2. Loss Overview



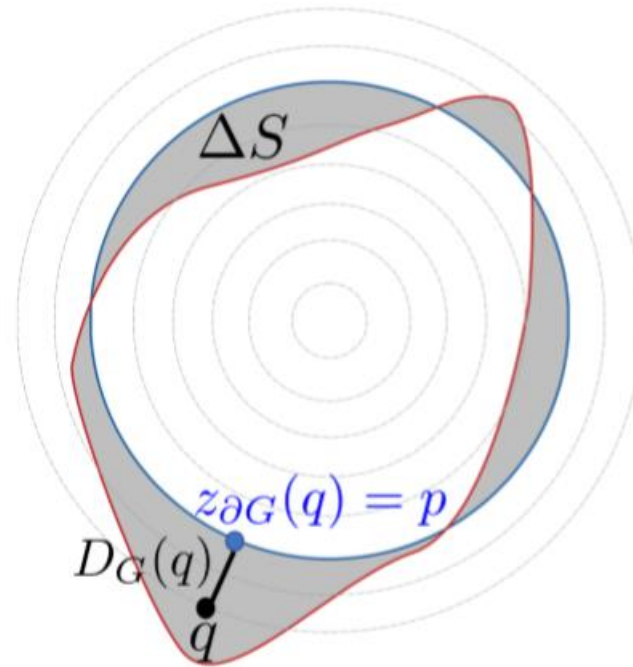
Relationship: Dice, Boundary, and HD Loss



Dice loss

$$\begin{aligned}
 &= 1 - \frac{2|S \cap G|}{|S| + |G|} \\
 &= \frac{|S| - |S \cap G| + |G| - |S \cap G|}{|S| + |G|} \\
 &= \frac{\Delta S}{|S| + |G|}
 \end{aligned}$$

$$\Delta S = (S \setminus G) \cup (G \setminus S)$$



Boundary loss

$$\begin{aligned}
 D_G(q) &= \|q - z_{\partial G}(q)\| \\
 \text{Dist}(\partial G, \partial S) &= 2 \int_{\Delta S} D_G(q) dq
 \end{aligned}$$

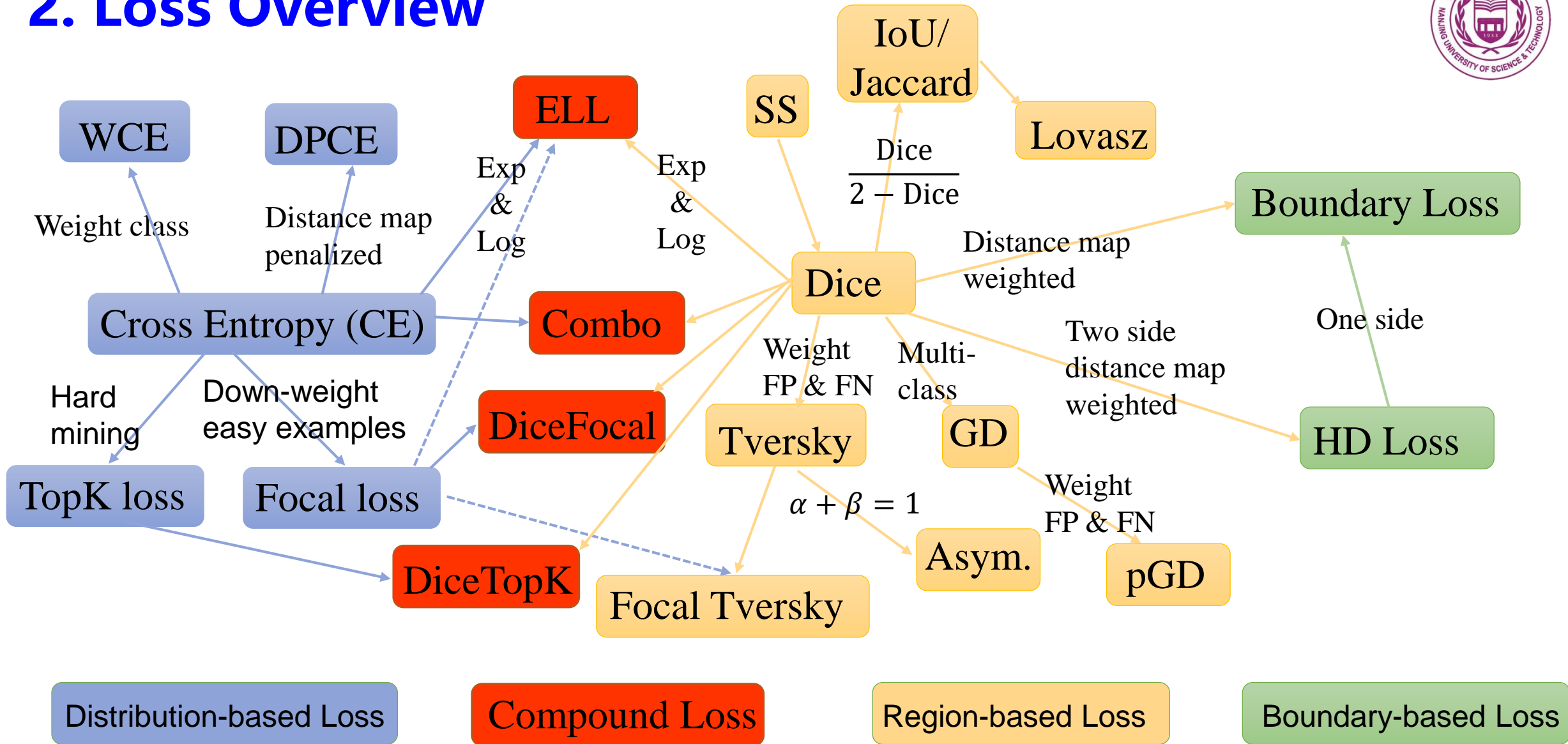
To some extent, all the three loss functions aim to minimize the **mismatch region**.

The key difference is the **weighting method**.

HD loss

$$= \frac{1}{|\Omega|} \sum_{\Omega} \Delta S \cdot (D_G + D_S)$$

2. Loss Overview





3. Code & Reference

Talk is cheap, here is the code (pytorch):

<https://github.com/JunMa11/SegLoss>

JunMa11 / SegLoss

Unwatch 17 Star 202 Fork 37

Code Issues 0 Pull requests 0 Projects 0 Wiki Security Insights Settings

A collection of loss functions for medical image segmentation Edit

Manage topics

45 commits 1 branch 0 releases 1 contributor

Branch: master New pull request Create new file Upload files Find file Clone or download

JunMa11 Add Non-Adjacency loss Latest commit 9bd70b8 on 2 Sep

losses_pytorch	Add files via upload	2 months ago
test	Add files via upload	2 months ago
README.md	Add Non-Adjacency loss	2 months ago



Thanks for Reading!